

(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau



(43) International Publication Date
25 October 2001 (25.10.2001)

PCT

(10) International Publication Number
WO 01/79449 A2

(51) International Patent Classification⁷: C12N Jose, CA 95117 (US). DRMANAC, Radoje, T. [YU/US]; 850 East Greenwich Place, Palo Alto, CA 94303 (US).

(21) International Application Number: PCT/US01/08656

(22) International Filing Date: 16 April 2001 (16.04.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/552,929 18 April 2000 (18.04.2000) US
09/770,160 26 January 2001 (26.01.2001) US

(63) Related by continuation (CON) or continuation-in-part (CIP) to earlier applications:
US 09/552,929 (CIP)
Filed on 18 April 2000 (18.04.2000)
US 09/770,160 (CIP)
Filed on 26 January 2001 (26.01.2001)

(71) Applicant (for all designated States except US): HYSEQ, INC. [US/US]; 670 Almanor Avenue, Sunnyvale, CA 94086 (US).

(72) Inventors; and

(75) Inventors/Applicants (for US only): TANG, Y., TOM [US/US]; 4230 Ranwick Court, San Jose, CA 95118 (US). LIU, Chenghua [CN/US]; 1125 Ranchero Way #14, San

(74) Agent: ELRIFI, Ivor, R.; Mintz, Levin, Cohn, Ferris, Glovksy and Popeo, P.C., One Financial Center, Boston, MA 02111 (US).

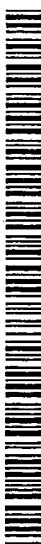
(81) Designated States (national): AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CO, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, US, UZ, VN, YU, ZA, ZW.

(84) Designated States (regional): ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:

— without international search report and to be republished upon receipt of that report

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.



WO 01/79449 A2

(54) Title: NOVEL NUCLEIC ACIDS AND POLYPEPTIDES

(57) Abstract: The present invention provides novel nucleic acids, novel polypeptide sequences encoded by these nucleic acids and uses thereof.

SEQ ID NO: of nucleotide sequence	SEQ ID NO: of peptide sequence	Method	SEQ ID NO: in USN 09/770,160	Nucleotide location of first codon for peptide sequence	Nucleotide location of last codon for last amino acid of peptide sequence	Amino acid sequence (X=Unknown; *=Stop codon; /=possible nucleotide deletion; \=possible nucleotide insertion)
						RKAASQSDKPAEKKEDESQMEDPS TSPSPGTRAASQPPNSSK\AGRKPW DRNNPLRNPLSSNLVRNPLLAGPR KLRAPFSQQPHSRMKPAGS\VSDMA LDAFDLDRMKQE*KEVVRELHK GERKEIID\AIRQEA\SGISRKKNLGH RAHPPTRTSFICSRPRLM
1746	7243	A	1876	1	668	GERGVARHDRPRGTLREYKVVGRC LPTPKC\HTPPL\YR\MRIFAP\SMSSL SPRFWYFVSQLKKDEESLQWRFSTY CAQVFEKSP\LRVK\NFGIWLRYD RSG\THNMY\REY\RDLDHPQAPVHP SCLTRDNGVAPAPAA\HEAHFHFRF ERLEEVAGQQDCRRPGCSKQFPRIS RFKFPAPPGLRQRDKP\RF\TTKRP KTFLKVQGPSSGVCPQNKTQETPR
1747	7244	A	1877	1	1059	
1748	7245	A	1878	87	260	
1749	7246	A	1879	1	1254	
1750	7247	A	1880	160	615	PSLNTYVTSPLSENFSARYRNHSND LTCVHTELQNKT\KLT\TVLEGDILDEP FLKRACQ\DVSV\NHTACI\DVFGVTT HRESIMNVNVKGRVAWGGDKARW GNEDQKEGQEGKRSLSIEHLLCSGP SDFADHYQLGELKAAIFSFIDEKTRT EQ
1751	7248	A	1881	53	1338	CPLQGHPRVTLES DLLPSIFCFLVSD SCYFGLATMGWSCLVTGAGGLLQQ RIVRLVVEEKELKEIRALDKA\FRPEL REEFSKLQNK\TKLT\TVLEGDILDEP LKS\ARD\RLRSI\HTACFHLM\MSFGV VTHREFF\VN\NVQC\KVPSSC*EACVQ ASVPVFIY\TSSIEVAGPNSYKE\I\QNG HEEEPLENTWPAPY\PRSKKLA\KKA VLAANGWNLK\NGGALYT\CALRPM YIYGEGRFLSVS\NEALNNNGILSS VGKFST\VN\PVYV\GNVAWGHILAL RALQDPKKAPSIRGQFYYISDDTPH QSYDNLNYTLSKE\FGPPPLDSRWS FPLSLMYW\IGFLLGNR*GFL\RPIY TYRPPFNRH\SSHCSN\ALFHLLFIKE GFSEILGVLRPLL\TAGGGKAKAGKR VGSWWV\VPFDPAQGRN\LEVPRIQ
1752	7249	A	1882	3	575	HS\LFGTSEVINKLLV\PDAMGHFTE DKAT\T\TSLWGK\VN\VEDAGGE\TP GKGSLVVY\WTQRF\FD\SG\NLSS ASA\MGKPPKSKAHG\KKVLT\FLGT MPTK\HLE*FSRGTF\CP\SLK*TCTC*Q ACMWDPGGTFKLPGENVAGLTVFG QSHFRQKN\TPEGAR\FFLGRKM\GD LELASALVPSRLPLKPLGP
1753	7250	A	1883	1	960	GRPAPE\DGGLSLPNAAMARGPKK HLKRVAA\PKHWMLDKLTGVFAPR PSTGPHKL\RECLPF\I\FLRNRLK\Y LTS\DEVKKICMQRFIK\DGQVR\TD IT\Y\AGFMDV\N\SIDKDGR\EF\SVL\Y LIDTQGVRFCL*HRJTP*GAKVQSC AKMRKILLWAPKGIPSSWVTHDAR

SEQ ID NO: of nucleotide sequence	SEQ ID NO: of peptide sequence	Method	SEQ ID NO: in USSN 09/770,160	Nucleotide location of first codon for peptide sequence	Nucleotide location of last codon for last amino acid of peptide sequence	Amino acid sequence (X=Unknown; *=Stop codon; /=possible nucleotide deletion; \=possible nucleotide insertion)
						NHPATPDPPSSKVN*YHFRLDLETG KDYLISSKFDTW*PCVMVTGGA\N LGRNWVLITN\RERHPGIF*PLVHV IDANGNLLATSDFSNIFWLLGKG KPW\ISL\PRKGKIPPHHLLERDKRL AAKQSSWVKWGPWVTWSDLLVP
1754	7251	A	1884	1	1218	FFQNSARGAGAGWQLPWTRFVWT SGLLEINETLVIQQRGVRIYDGEEKI KFDAGTLLSTHRLIWRDQKNHEC CMAILLSQ\VFIEEQAVAGIGKSAKI VVHL\HPAPPNKEPGPVFQSSKNSYI KLSFKEHGQIEFYRRLSEEMTQRRW ENMPVSQSLQTNRGPQPGRIRAVGI VGTERKLEEKRKETDKNISEAFEDL SKLMIKAKEMVELSKSIANKIKDKQ GDITEDETIRFKSYL\LSMGIANPVT RETYGSGTQYHM\QLAKQL\AWNIA RVPLEERGGIMSLTEVYCLVNRARG MELLSPEDLVNACKMLEALKPLR LRVFDSGVMVIELQSHKEEEMVAS ALETVSEMGSLTS*EFAKLVGMCSV LAKERLLAEMKGHLCRDDSVEGL RFYPNLFMTQS
1755	7252	C	1885	179	361	MPKVCVHNFLKTSSERDLFALMN TVGKKHSIMSEKGRSKFLHLIDSK KNEDPHLDGT*
1756	7253	A	1886	2	913	RRLLLFGWARSGAVSLGSAGVSSS GFLTAPHSRRLTAAAAAAAGGAWRF EAERHRGWAEEEQQPEGGAVCPG TERPCAMAYAYLFKYIIIGDTGGGR SCLLLQFTDKRFQPSAMTLTNGVEF GARMITIDGKQIK\LQIWDTAGQES FRSITRISYY\RGAAAGALLVYDITR DTSTHLLTWLEDA\RHSHFQHG LCLLGKNSDL\ESRKE\VSKKRKEGE SFLQPRNHGLHLPWKTSCKNCFPM* KEAFINTSKRNFIEKIQEGVFDINNE A\NGIKIGP\QHAATNA\THAG\NQGG QQAGGGCC
1757	7254	A	1893	138	426	FIHSHCCIVFRLFIHFSLHPKVIHSPIN SLLRIFQF*AIMNSTV*NLIHVFW*V YTFPGINPKKGIARL*GVYIFSFITY CQTVFQSDCKKAPF
1758	7255	A	1894	45	1057	FLVFLVETGFHHVAQAVLELLASSD PPALAPPKWCWDYRCCELLRLAEFCFL RTEFWYLLFFFFWRRSLALSPRLLEC SGANL\THCNLR/LPGFKQFSCLSLSS SWDYRCMPHHLATFFVF/SVETGFH RVAQASLELLSSGSLPALA/FPKCW DYRAKATV/WPSPGVSSFILGL*TS* FHSLEPYLHAWKTTSHLPTKEALT W\VSHTAKTKHLWILVSI\MEF*VA LIS/SFFLGP\GGK*T*VTAPQCPSLGQ DTLS*FLHA\ACTRSVPY\PGLA/CGPS LWLTRVLLPTPP*QQHNP/DTLEKT SFPGPHWIL*TPQPSLSETPAPKVPP FPAFGSIPTHEEPGLP
1759	7256	A	1895	2	289	

SEQ ID NO: of nucleotide sequence	SEQ ID NO: of peptide sequence	Method	SEQ ID NO: in USN 09/770,160	Nucleotide location of first codon for peptide sequence	Nucleotide location of last codon for last amino acid of peptide sequence	Amino acid sequence (X=Unknown; *=Stop codon; \=possible nucleotide deletion; \=possible nucleotide insertion)
1760	7257	A	1896	1	397	
1761	7258	A	1897	1	410	STMISPVLILFSSFLCHVAIAGRTCPK PDDLPFSTVPLKTFYEPGEEITYSC KPGYVSRGGM\RKFICPLTGLWPIN TLKCTPRVCVPFAGNLRKMGAVRLIT DFLNYSPTRFSFSLLTWGFILEWAL DS\AKCIEGG
1762	7259	A	1898	19	1215	CQCDSSTMFSRCSSLFSSFLCHVAI AGRTCPKPDDLPFSTVPLKTFYEP G\EEITYSCKPGYVSRGGIEESLSCPL \TGTVGPFNTSGNVTPRVC\PF\AGIFR KMGGRTLITF*NYPNTDPVFSLLTL GF*FWNGALDFWP\SC\TGGKGK\W\ S\PELPGLVAPI\CPP\PSIP/TGFATLH VLLRPFRLGNNSPIGDTAVFECLA HNMAMFG\NDTITCTTHGKLDLNY PECRGSKMPPFPHQDPDNGI\W*TYP CQNPNTL\TRVKA\PHLGLPHDGIFS GMGPRKE\I\EC*PQTW\KPGS\WPLA PSW*KPSLVKGTPVKKRPTVV/YPQ GERVKDSREKFKEWECLHG**KFLS FCKNKEKKCSYTEDAQCIDGTIEVP KCFK\EHSSLAFWKT\DA\SVKPC
1763	7260	A	1899	58	446	
1764	7261	A	1900	1	954	MGEVSGTS\DC\TDDQCRQVKALEG GKAARGHRSKJ\KIRFFRPGGLGP\GP AITA\AGMP\RVYIGRLSYQAREHA \VERLLNGHAKILEVDLKNGYGFVE FDDLRDADDAVYELNGKDLCGERV IVEHARGPRRDGSYGS\GRSGYGYR RSGRD\K\YGP\TRTEDR\I\VE\N\LT\SR CSW\QDL\DY\MRQAGEV\TYADAHK GRQKMKGVIEFVSYSDM\KRALEKL DGTEVNGR\KIRLVEDKPGSRRRRSY SRSR\SHSRSRSRSRHSR\KRSRSRSGSS KSSHSK\RSRSRSRSGSRSRSK\RSRSR\Q SRSRSKKEKRSRSPSKDKS\RSRSR\HSA\ GKSRSKSKDQAE\EKFQ\NNDN\VGK PKSRSR\SHKSKS\RSRSR\Q\ER\VEE GRKRG\SF*Q\Q\EAQEKSLRQ\RSR\ SR\SKAG\SR*PVDR\RSKSKDKR\K\SR KRSREESRSRSRSR\KRSRSR\KRG\ S KRD\SKAS\SKKKKED\TDR\Q\RSR\ PSRSV\SKEREHA\RS\LESS\Q\REG\RG ESENAGTNQ\EDPGPG\PRSM\SK\SKP N\LP\RM\HRSKJ\KSQASKTP\ISGP\MSR SR\SA\RSRSP\RSRSK\RSRSR\Q\RSR\ KKEKRSRSPSKDKSLQPQP
1765	7262	A	1901	3	180	
1766	7263	A	1902	227	440	GMHNV\CV\AVNE*FCGFIIR*SLAE RRQ\IS*EFQL\KFTLC\LE\I\ARRAC RESMA\SPVAGSW\SHF\PEREF
1767	7264	A	1903	2	438	HEELDT\SERKIEFDSASGTY\TLYLN\ GDAHFE\EP\Q\SL\W\NV\ADL\VHQ\SP\PE EKAPLDLSCP\QNL\FTPK\Q\EI\Q\W\IRI GAI\N\VS\N\FT\A\PI\STI\FT\H\LG\H\A\AM LGL\MY\Y\W\T\Q\LN\MF\Q\T\K\Y\LA\L GS\T\FL\AG\N\R\ML\A\QQ\A\K\RT\AH

SEQ ID NO: of nucleo-tide sequence	SEQ ID NO: of peptide sequence	Method	SEQ ID NO: in USSN 09/770,160	Nucleotide location of first codon for peptide sequence	Nucleotide location of last codon for last amino acid of peptide sequence	Amino acid sequence (X=Unknown; *=Stop codon; /=possible nucleotide deletion; \=possible nucleotide insertion)
1768	7265	A	1904	1	1660	
1769	7266	A	1905	156	2369	PVLKTHPGPQSLPRVPGVPCGGLLE PLSRAEVSPRGRLRDLGGMAPPG SSTVFLLALTIIASTWALTPTHYLTK HDVERLKASLDRPFTNLESAFYISIV GLSSLGAQVPDAKKACTYIIRSNLDP SNVDSLKYVA\QA\SQGLSGCEISIN ETKDLLLA\AVSEDSSVYPRSYHAS WQL*SGLLGSLWAVPKESTQVAL NWLVFKQGKETVL\ATVQALQTAS HLSQQADLRSIVEE\EDLVARLDEL GGLYLQ\FEEGLETTALFVAATYKA /LMDHVGT\EP\SIKE\ADQV\IQLMNAI F\SKKNFES\SEAFS\ASG\AAVLS HNRYHVPVVVVPEGSASDTHEQAI LRLQVTNVLSQPLTQATVKLEHAK SVASRATV\QKTSFTP\VGIVFELNF MNVKFSGG*CDF\VEVEGDNRYIS\N TVELRVQDPPTEVGIT\NVDLSTV\DKDQSIAP\QTRTVTYPAKAGTFH SAGQATRNFGVLSSW*DVENTG\AE LTPHQTFVRLHNQKTGP\SGCLFAE PGQQGTCYK\ELDTSERKGLNLT\SR SGTYTLYLIIG*CQL*RTQILWKCGL MWV\KFP\GKEASFDC\LC\CSQEP\FS PKQGNFRHLFPG\RP\GRRAP\PPWCP NTFTAPESFFGPLL/LCFLRLLWIRD WVPKCLPTFTCFLSTIIFHPWDML AYAGTSMYVY*TQAQPCSQ\TLEVP W\PILGQCDRFLAGQSGMLAPARQV KRIAEEQSSRLAKYRTLRTAH
1770	7267	A	1906	37	404	PQLSRCRSEC\MYVNPTVVMTSMGQ ATWSDPHKAKTMLNR\IPLGKFAGE SGGSPASVVA\VPV\CALGRGGRER WAAASFLYAPDPRPAH\VE\HVN A\ILFL\LLSDRSGMTTGSTLP\VEGGFW AC
1771	7268	A	1907	271	1086	YTQCPGIEPV\CV\DLG\DW\EA\TERAL GSVGP\DL\VVNNA\AV\LLQPF\LEV TKEAFDR*ACEGGGTSGRGC\PGGRS SPNL*PGSVP\PLDPLRVNLRA\VIQV SQIV\A\GLI\A\RGVPTGPS*NVSSQC FPAGQ*TNHSV\LLPTKGVPLDM\LD QG*WAL\ELGPHKLSRC\SGVNA\W NPHSGG*RSMGP\GPPWSDPHK\AKI MLN\IPLGK\FA\GE\SE\VE\HVN\AIL FL\LSDRSGMTTGSTLP\VEGGFW\AW LSSL\HTPQAPWAC\ILT\NP\SNKT
1772	7269	A	1908	2	305	ARESGSLVAPRSRPPWEHGLPGEHS *DAPRPHKSPTLPWLPHLHLSKEAL DTHQRSQHE\ECMPLYKFTPTSEKR PQLMLPLPEQQCEQLCRFGSTPVTA A
1773	7270	A	1909	2	529	GTVAACGACYWLLGLMA\VRASFE NNCEIGCFAKLTNTYCLVA\IGGSEN FYSV\FEGELS\TIPV\H\AS\IAGCRIIG RM\CVG\TE\ELIADVLK\VEVFRQTV A\DV\LVGSYCVFSNQGGLV\HPKTSIE

WHAT IS CLAIMED IS:

1. An isolated polynucleotide comprising a nucleotide sequence selected from the group consisting of SEQ ID NO: 1-5497, a mature protein coding portion of SEQ ID NO: 1-5497, an active domain of SEQ ID NO: 1-5497, and complementary sequences thereof.
5
2. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide hybridizes to the polynucleotide of claim 1 under stringent hybridization conditions.
- 10 3. An isolated polynucleotide encoding a polypeptide with biological activity, wherein said polynucleotide has greater than about 90% sequence identity with the polynucleotide of claim 1.
4. The polynucleotide of claim 1 wherein said polynucleotide is DNA.
- 15 5. An isolated polynucleotide of claim 1 wherein said polynucleotide comprises the complementary sequences.
6. A vector comprising the polynucleotide of claim 1.
- 20 7. An expression vector comprising the polynucleotide of claim 1.
8. A host cell genetically engineered to comprise the polynucleotide of claim 1.
- 25 9. A host cell genetically engineered to comprise the polynucleotide of claim 1 operatively associated with a regulatory sequence that modulates expression of the polynucleotide in the host cell.
10. An isolated polypeptide, wherein the polypeptide is selected from the group consisting of:
 - (a) a polypeptide encoded by any one of the polynucleotides of claim 1; and
 - 30 (b) a polypeptide encoded by a polynucleotide hybridizing under stringent conditions with any one of SEQ ID NO: 1-5497.
11. A composition comprising the polypeptide of claim 10 and a carrier.
- 35 12. An antibody directed against the polypeptide of claim 10.

13. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polynucleotide of claim 1 for a period sufficient to form the complex; and

5 b) detecting the complex, so that if a complex is detected, the polynucleotide of claim 1 is detected.

14. A method for detecting the polynucleotide of claim 1 in a sample, comprising:

a) contacting the sample under stringent hybridization conditions with 10 nucleic acid primers that anneal to the polynucleotide of claim 1 under such conditions;

b) amplifying a product comprising at least a portion of the polynucleotide of claim 1; and

15 c) detecting said product and thereby the polynucleotide of claim 1 in the sample.

15. The method of claim 14, wherein the polynucleotide is an RNA molecule and the method further comprises reverse transcribing an annealed RNA molecule into a cDNA polynucleotide.

20 16. A method for detecting the polypeptide of claim 10 in a sample, comprising:

a) contacting the sample with a compound that binds to and forms a complex with the polypeptide under conditions and for a period sufficient to form the complex; and

b) detecting formation of the complex, so that if a complex formation is detected, the polypeptide of claim 10 is detected.

25

17. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

a) contacting the compound with the polypeptide of claim 10 under conditions sufficient to form a polypeptide/compound complex; and

30 b) detecting the complex, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

18. A method for identifying a compound that binds to the polypeptide of claim 10, comprising:

- a) contacting the compound with the polypeptide of claim 10, in a cell, under conditions sufficient to form a polypeptide/compound complex, wherein the complex drives expression of a reporter gene sequence in the cell; and
- 5 b) detecting the complex by detecting reporter gene sequence expression, so that if the polypeptide/compound complex is detected, a compound that binds to the polypeptide of claim 10 is identified.

19. A method of producing the polypeptide of claim 10, comprising,
 - 10 a) culturing a host cell comprising a polynucleotide sequence selected from the group consisting of a polynucleotide sequence of SEQ ID NO: 1-5497, a mature protein coding portion of SEQ ID NO: 1-5497, an active domain of SEQ ID NO: 1-5497, complementary sequences thereof and a polynucleotide sequence hybridizing under stringent conditions to SEQ ID NO: 1-5497, under conditions sufficient to express the polypeptide in said cell; and
 - 15 b) isolating the polypeptide from the cell culture or cells of step (a).

20. An isolated polypeptide comprising an amino acid sequence selected from the group consisting of SEQ ID NO: 5498-10994, the mature protein portion thereof, or the active domain thereof.

- 20 21. The polypeptide of claim 20 wherein the polypeptide is provided on a polypeptide array.

22. A collection of polynucleotides, wherein the collection comprises the sequence information of at least one of SEQ ID NO: 1-5497.
- 25 23. The collection of claim 22, wherein the collection is provided on a nucleic acid array.

24. The collection of claim 23, wherein the array detects full-matches to any one of the polynucleotides in the collection.

- 30 25. The collection of claim 23, wherein the array detects mismatches to any one of the polynucleotides in the collection.

26. The collection of claim 22, wherein the collection is provided in a computer-readable format.

27. A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising a polypeptide of claim 10 or 20 and a pharmaceutically acceptable carrier.

5 28 A method of treatment comprising administering to a mammalian subject in need thereof a therapeutic amount of a composition comprising an antibody that specifically binds to a polypeptide of claim 10 or 20 and a pharmaceutically acceptable carrier.